

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 2000-330582

(43)Date of publication of application : 30.11.2000

(51)Int.Cl. G10L 13/00
G10L 11/00

(21)Application number : 11-137123

(71)Applicant : NIPPON TELEGR & TELEPH CORP
<NTT>

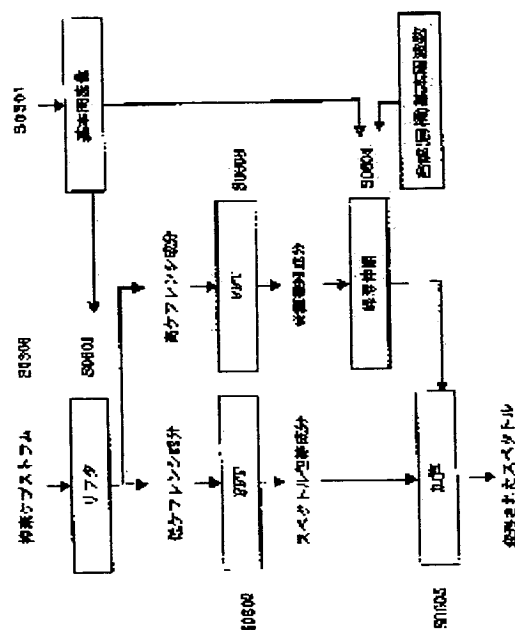
(22)Date of filing : 18.05.1999

(72)Inventor : TAKANO SATORU
ABE MASANOBU(54) SPEECH TRANSFORMATION METHOD, DEVICE THEREFOR, AND PROGRAM
RECORDING MEDIUM

(57)Abstract:

PROBLEM TO BE SOLVED: To provide a speech transformation method in which fine components are maintained even if a fundamental frequency is transformed, by considering that the small components of an amplitude spectrum before flattening have an influence on sound quality.

SOLUTION: According to this speech transformation method, a transformed speech is obtained by processing an input speech signal by FFT by multiplying it by a complementary Gaussian window with a three-pitch length of the fundamental frequency; obtaining a logarithmic amplitude spectrum therefrom; processing the spectrum by inverse FFT to obtain a complex cepstrum; dividing it into low quefrency components and high quefrency components; processing each of them by FFT (S0602, S0603) to obtain a spectrum envelope component and sound source fine components; expanding and contracting the sound source fine components in the frequency domain with a ratio of the fundamental frequency to a target fundamental frequency (S0604); summing this and the spectrum envelope component to obtain a transformed spectrum; and converting it into the time domain.



BEST AVAILABLE COPY

LEGAL STATUS

[Date of request for examination] 08.11.2001

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number] 3557124

THIS PAGE BLANK (USPTO)

[Date of registration] 21.05.2004

[Number of appeal against examiner's decision
of rejection]

[Date of requesting appeal against examiner's
decision of rejection]

[Date of extinction of right]

THIS PAGE BLANK (USPTO)

(19) 日本国特許庁 (J P)

(12) 公 開 特 許 公 報 (A)

(11) 特許出願公開番号
特開2000-330582
(P2000-330582A)

(43) 公開日 平成12年11月30日 (2000. 11. 30)

(51) Int.Cl. ⁷	識別記号	F I	テマコード* (参考)
G 1 0 L 13/00		G 1 0 L 7/02	D
11/00		9/16	

審査請求 未請求 請求項の数 5 O L (全 10 頁)

(21) 出願番号 特願平11-137123

(22) 出願日 平成11年5月18日 (1999. 5. 18)

(71) 出願人 000004226

日本電信電話株式会社
東京都千代田区大手町二丁目3番1号

(72) 発明者 ▲高▼野 哲

東京都新宿区西新宿三丁目19番2号 日本
電信電話株式会社内

(72) 発明者 阿部 匡伸

東京都新宿区西新宿三丁目19番2号 日本
電信電話株式会社内

(74) 代理人 100066153

弁理士 草野 卓 (外1名)

(54) 【発明の名称】 音声変形方法、その装置、及びプログラム記録媒体

(57) 【要約】

【課題】 規則合成における基本周波数変形を高品質で可能とする。

【解決手段】 入力音声信号に対し、その基本周波数の3ピッチ分の長さで相補的ガウス窓を掛けてFFTし、その対数振幅スペクトルを求め、これを逆FFTして複素ケプストラムを得、これを低ケフレンシ成分と高ケフレンシ成分に分け、それぞれをFFTして (S0602, S0603)、スペクトル包絡成分と、音源微細成分を得、基本周波数と目標基本周波数比で音源微細成分を周波数領域で伸縮し (S0604)、これとスペクトル包絡成分とを加算して変形されたスペクトルを得、これを時間領域に変換して変形音声を得る。

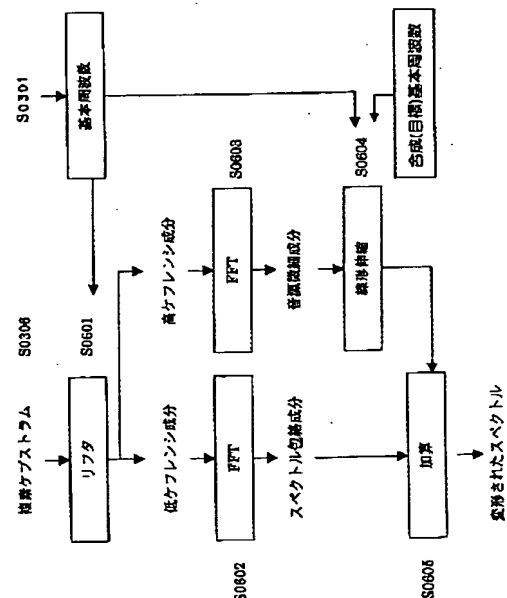


図 5

【特許請求の範囲】

【請求項 1】 入力音声信号に時間窓を乗じる過程と、その時間窓を乗じた入力音声信号を周波数領域に変換して短時間スペクトルを分析する過程と、前記スペクトルからスペクトル包絡と微細構造を分離する過程と、前記入力音声信号から基本周波数を分析する過程と、前記基本周波数と目標基本周波数との比に比例して前記微細構造を周波数領域で伸縮する過程と、前記スペクトル包絡と前記伸縮された微細構造とを加算してスペクトルを合成する過程と、前記合成されたスペクトルを時間領域の信号に変換する過程と、を有する音声変形方法。

【請求項 2】 前記基本周波数に反比例するように前記時間窓の窓長を決定する過程を有し、前記時間窓を乗じる過程は、前記時間窓として前記決定された窓長の相補的ガウス窓関数を乗じる過程であることを特徴とする請求項 1 記載の音声変形方法。

【請求項 3】 入力音声信号に時間窓を乗じる時間窓乗算部と、その時間窓が乗算された入力音声信号を周波数領域に変換して短時間スペクトルを分析するスペクトル分析部と、前記スペクトルからスペクトル包絡と微細構造を分離するスペクトル分離部と、前記入力音声信号から基本周波数を分析する基本周波数分析部と、前記基本周波数と目標基本周波数との比に比例して前記微細構造を周波数領域で伸縮する微細構造伸縮部と、前記スペクトル包絡と前記伸縮された微細構造とを加算してスペクトルを合成するスペクトル合成部と、前記合成されたスペクトルを時間領域の信号に変換する信号合成部と、を有する音声変形装置。

【請求項 4】 前記基本周波数に反比例するように前記時間窓の窓長を決定する窓長設定部を有し、前記時間窓乗算部では、前記時間窓として前記決定された窓長の相補的ガウス窓関数が用いられる、ことを特徴とする請求項 3 記載の音声変形装置。

【請求項 5】 請求項 1 又は 2 の何れか 1 項に記載の音声変形方法をコンピュータが実行するプログラムを記録した記録媒体。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】この発明は、テキストからの音声合成をはじめとする規則音声合成技術において、合成目標の韻律に合わせるために音声素片を基本周波数変形する際に生じる合成音声の品質劣化を抑制したり、自然音声进行分析合成して声質変形する際に生じる品質劣化を抑制することを目的とする音声合成方法及びプログラム記録媒体に関するものである。

【0002】

【従来の技術】最近主流となっている PSOLA 合成法をはじめとする波形合成方式は、音声波形をそのまま使うことから、変形量が少ない際に高音質な反面、変形量が大きいと歪みが目立つ。高音質を目指すためには、変形量を少なくするために大量の素片を必要としてしまう。また、自然な響きと個性を保持することを長所とする CHATR 法は大量の音声データを持ち、その中から出力する文に応じて適合する音韻を選び出す方式で、高音質な反面、信号処理を行わないので一般に所望の基本周波数パターンで音声合成することは難しい。

【0003】柔軟な変形を行うためには、音声进行分析してモデル化するボコーダ型の分析合成方式が有利であるが、従来のボコーダは波形合成に対して音質が劣るといのが定説である。ボコーダの中でも短時間スペクトル分析をもとにした合成法では、ハミング窓にもとづく短時間スペクトルを平滑化したものをもとに合成しているのがほとんどである。

【0004】

【発明が解決しようとする課題】人間の発声では、声帯と声道はつながっており、基本周波数の調節に伴って声道形状にいくらかの変化が生じることから、基本周波数の変化とともにスペクトルの構造も変化する。そこで、音声の基本周波数変形には基本周波数の変化量に応じて、スペクトル形状を変換することも必要になる。しかし、従来のボコーダ型合成法は声帯音源情報と声道形状にあたるスペクトル包絡情報を分離してモデル化しており、異なる基本周波数パターンを持っている変換対象の素片のスペクトルをそのまま用いて合成するのがほとんどであった。その場合、合成する基本周波数とスペクトルの不整合を防ぐために、スペクトル包絡は振幅スペクトルを平滑化したものを用いていた。

【0005】この発明では平滑化する前の振幅スペクトルの微細成分が音質に影響をおよぼしていると考え、基本周波数を変形しても微細成分を保持するような音声変形方法を提供する。

【0006】

【課題を解決するための手段】この発明によれば、入力音声信号に時間窓を乗じ、その時間窓を乗じた音声信号を周波数領域に変換して短時間スペクトルを分析し、そのスペクトルからスペクトル包絡と微細構造を分離し、また入力音声信号から基本周波数を抽出し、その基本周波数と目標基本周波数との比に比例して微細構造を周波数領域で伸縮し、その伸縮された微細構造とスペクトル包絡を加算してスペクトル合成し、その合成されたスペクトルを時間領域に変換する。

【0007】基本周波数に反比例するように時間窓の窓長を決定し、その窓長の相補的ガウス窓関数を時間窓として用いる。

(1) 音声の一般的性質として、スペクトル包絡は声道

の特性を反映し、微細構造は基本周波数とその高調波成分を示す。基本周波数が増加する場合には、スペクトル包絡は変化せず、微細構造が基本周波数に応じて変化する。

(2) 本発明によれば、目標周波数に音声を変形しても、スペクトル包絡は変化せず、微細構造が基本周波数に応じて変化する。したがって、変形された音声を聴取しても自然性を維持できる。

(3) さらに、音声を切出すときに用いる時間窓として音声の基本周波数の周期に比例した窓長をもつガウス相補型時間窓を用いることによって、基本周波数成分を反映した微細構造の抽出が損なわれない。そのため、この発明により目標基本周波数が得られるように基本周波数を変化させて音声を変形してもその自然性が確保できる。

【0008】短時間スペクトルを求める場合、フーリエ変換の前に窓関数をかけて音声波形を切り出すが、従来の合成法では周波数分解能が高いハミング窓を用いるものがほとんどであった。この発明の実施例では、相補的なガウス窓を用いた短時間スペクトルをもとに分析することが特徴である。ハミング窓による短時間スペクトルと相補的なガウス窓（文献1参照）による短時間スペクトルを図9に示す。比較のためにハミング窓による短時間スペクトルを-1 dB 下方にシフトしている。スペクトルの包絡は似ているが、微細成分の上下動が大幅に異なることがわかる。相補的なガウス窓による短時間スペクトルはスペクトル包絡をとってももとのスペクトルに近い一方、微細な成分もある程度併せ持っている。

【0009】

【発明の実施の形態】図1にこの発明の基本構成を示す。入力音声はステップS0201のスペクトル分析部で短時間スペクトル分析および基本周波数抽出をされ、それをもとにステップS0202のスペクトル変形部で入力音声と目標合成音声の基本周波数の違いに応じてスペクトル変形を行い、変形されたスペクトルに基づいてステップS0203の重複加算合成部で重複加算法により合成音声を生成する。

【0010】以下、この発明を規則音声合成に適用する場合の実施例を述べる。図1中のステップS0201にあたる短時間スペクトル分析の処理例を図2に示す。入力音声は図3に示すように固定フレーム間隔 ΔT で、つまり ΔT づつ順次シフトさせながら分析され、 n フレーム目の分析結果の短時間スペクトル $X(\omega, n)$ はフレームの中心の時間位置 $n \Delta t$ と結びつけたパラメータとなる。まず、ステップS0301で入力音声については各フレーム中心点 $n \Delta t$ での基本周波数を抽出する。実施例ではTEMPO法（文献1参照）にもとづいて基本周波数抽出に加えて有声無声判定も行う。各フレームについて、フーリエ変換の次数を N とすると、ステップS0302で音声波形からフレーム中心の前後 $N/2$ 点

を切り出し、ステップS0303では当該フレームのステップS0301で抽出した基本周波数に基づき、3ピッチ分の長さの相補的なガウス窓をかけてステップS0304でFFTにより短時間スペクトルを求める。この窓かけの時間長、つまり短時間スペクトルを求める時間長は1~5ピッチ分程度が好ましい。5ピッチより長くすると、不要な波形が加わって平均化され、微細成分を取り出すことが困難になる。好ましくは3ピッチ分程度がよい。ステップS0305では求めた短時間スペクトルの絶対値の対数を取り、対数振幅スペクトルを得る。ステップS0306では対数振幅スペクトルの逆フーリエ変換を行い、複素ケプストラム $C(\omega, n)$ を得る。

【0011】従来法の例としてSTRAIGHT法では図4に示すように、ステップS0304の振幅スペクトルに対してステップS0501で平滑化を行い、音源成分にあたる微細成分を除去している。この発明では微細成分を除去せず、保持した状態でスペクトル変形を行う。図1中のステップS0202にあたるスペクトル変形の処理例を図5に示す。ステップS0201で求めた入力音声の基本周波数をもとにケプストラムリフタを構成し、図2で求めた複素ケプストラム $C(\omega, n)$ に対して、ステップS0601で低ケフレンシ成分と高ケフレンシ成分に分離する。この分離は例えば基本周波数と対応する点よりわずかに(20~30程度(FFTが1024点の場合))低い点を境界として分けられればよい。低ケフレンシ成分はスペクトル包絡のケプストラムであり、この低ケフレンシに対しステップS0602でFFTを行うことによりスペクトル包絡 $E(\omega)$ が得られる。高ケフレンシ成分は音源成分のケプストラムであり、この高ケフレンシ成分に対し、ステップS0603でFFTを行うことにより音源成分であるスペクトルの微細成分 $R(\omega)$ が得られる。これらスペクトル包絡 $E(\omega)$ とスペクトルの微細成分 $R(\omega)$ の例を図6に示す。

【0012】音源成分 $R(\omega)$ については、微細成分保持および合成基本周波数との整合をとるためにステップS0604で入力音声の基本周波数と合成音声の基本周波数の比 r をもとに線形伸縮を行う。この $R(\omega)$ に対する線形伸縮を、この実施例では図7に示す。 $R(\omega)$ は離散フーリエ変換をもとにした離散データであるために、 $R(n)$ (n は整数)と表すことができ、周波数方向に伸縮すると離散値のサンプリング点からはずれる。そのため、伸縮後の微細成分のスペクトル $R'(m)$ (m は整数)は図7に示すように $r n \leq m < r(n+1)$ となる n に対して $R'(m) = \{ (r(n+1) - m) R(r n) + (m - r n) R(r(n+1)) \} / r$ により線形補間により求める。また、伸縮を行うために、基本周波数を高くする場合($r > 1$)には高域において有効周波数をはみ出す部分があるため、この部分を捨て去る。逆に基本周波数を低くする場合($r < 1$)に

は高域のスペクトルを作る必要がある。この実施例では足りなくなった周波数でそこを中心に線対称となるスペクトルを用いる（スペクトルの折り返し）ことにより高域を作成する。ステップS0605においてスペクトル包絡と変形された音源成分の加算を行い、合成基本周波数にあったスペクトルを再構成する。

【0013】音源成分 $R(\omega)$ に対する線形伸縮の他の例を以下に示す。

(1) 伸縮前の微細構造のスペクトル $R(m)$ ($0 \leq m < N/2$ 、 m は整数、 N はフレーム長)をフーリエ展開 (好ましくはFFT)して展開係数 $\rho(k)$ を得る ($0 \leq k < N/2$ 、 k は整数、 N はフレーム長)。ここで、 $\rho(k) = \sum_{m=0}^{N/2-1} R(m) \exp(2\pi jmk/N)$ (j は虚数単位)と算出する。また、 $R(m) = 2/N \sum_{k=0}^{N/2-1} \rho(k) \exp(2\pi jmk/N)$ という関係がある。

(2) 基本周波数 F_0 と目標基本周波数 F_0' とから伸縮後の微細構造のスペクトル $R'(m)$ を基底関数 \exp における変数 x を (F_0/F_0') 倍、つまり $(F_0/F_0')x$ と置換して展開する。

【0014】 $R'(m) = 2/N \sum_{k=0}^{N/2-1} \rho(k) \exp(2\pi jmk/N \times (F_0/F_0'))$

このような演算により、周波数幅 F_0 内における伸縮前の微細構造のスペクトル $R(m)$ における成分は周波数幅 F_0' に伸縮される。周波数は離散的に与えられるため、上記の線形補間によれば顕著に平滑化されたり、 $F_0' < F_0$ のときに $f/2$ (f は標本化周波数)を定義できないという問題が生じる。しかし、この方法によればかかる問題を生じず、一義的に解を与えることができる。

【0015】図1中のステップS0203にあたる重複加算合成部は図8に示す。この合成部はSTRAIGHT法に由来するものであり、詳細は文献1を参照のこと。ステップS0605で再構成されたスペクトルをもとにステップS0901で対数スペクトルを求め、ステップS0902でフーリエ逆変換により複素ケプストラムを求め、ステップS0903で最小位相化したスペクトルを求める。ステップS0904では有声音成分としてそのスペクトルに、位相操作として周波数領域でオールパスフィルタをかけて、ステップS0905でフーリエ逆変換により、インパルス応答を求めて、ステップS0906では合成する基本周波数の逆数に当たるピッチ間隔で重複加算する。また、ステップS0907では無声音成分としてそのスペクトルのインパルス応答を求め、ステップS0908で乱数列をたたみこんだものをステップS0909で固定間隔で重複加算する。ステップS0910では有声音成分と無声音成分をステップS0201で求めた有声音判定をもとに混合して合成音声を得る。

【0016】次に実験例を述べる。不均一な基本周波数

F_0 の変形が連続する規則合成音についてプリファレンステストで評価した。対象とする音声はそれぞれ1~2秒程度の長さのもの3種類である。CV-VC素片で合成対象を覆うように素片を用いた。 F_0 の変形量が大きく、単位が短い場合となるように選んだ、合成法としてPSOLA、STRAIGHT、本発明方法の3種類について比較した変形対象の F_0 を平行移動することにより全体的に高くする場合と、全体的に低くする場合でどう変わるか検討した。 F_0 の平行移動はもとの F_0 の1.2倍、1.4倍 (F_0 を高くする)、 $1/1.2$ 倍、 $1/1.4$ 倍 (F_0 を低くする)の4通りである。被験者は9人で3種類の合成音のうち2個を1組とした刺激音をランダムな順番で提示した。

【0017】その結果を図10に示す。上から順に3組ずつ、もとの合成 F_0 。パターンと4種類の平行移動した F_0 。パターンでの結果である。各横バーの中の数字は同一テキストについての左の合成音と右の合成音を比較したときに左の合成音を選んだ比率を示す。例えば1番上の横バーでは同一テキストに対し、左のPSOLAによる合成音の方が、STRAIGHTによる合成音より音質がよいとした率が38.9%であることを示す。図中の◎は危険率が1%、○は危険率5%で有意な差を示す。この結果から、 F_0 を低くする変形でPSOLAよりSTRAIGHTや本発明方法が良い評価を得ている。 F_0 を高くする変形ではこの発明方法とSTRAIGHTを比較するとこの発明方法の方が選択され、しかも◎、○印のものについてはこの発明方法によれば従来法より高品質が得られることの信頼性が高いことが理解される。

【0018】

【発明の効果】以上述べたようにこの発明によれば、入力音声のスペクトルをスペクトル包絡と微細成分（構造）とに分離し、その微細成分について目標基本周波数と原基本周波数との比に応じて伸縮させた後、スペクトル包絡と合成し、この合成スペクトルを時間領域に戻すことにより、スペクトル包絡は変化せず、微細成分のみを基本周波数に応じて変化させることができ、変形された音声を聴取しても自然性が維持される。

【0019】特にガウス相補型時間窓を用いると、微細成分（構造）の抽出が良好に行われ、変形音声の自然性がより良好に確保できる。この発明を分析合成音声および、規則音声合成に適用したところ、先にも示したように従来法に比べて高音質な合成音を得られることが確認できた。

【参考文献】

1. 河原「聴覚の情景分析と高音質音声分析合成法STRAIGHT」音講論、1-2-1、pp189-193、1997(9)

【図面の簡単な説明】

【図1】この発明の基本構成を示す図。

【図2】図1中の音声分析部の処理手順の例を示す図。

【図3】音声信号に対する分析フレームと時間窓との関係例を示す図。

【図4】従来の音声分析の一部を示す図。

【図5】図1中のスペクトル変形部の処理手順の例を示す図。

【図6】音源成分（微細成分）とスペクトル包絡の分離例を示す図。

*

* 【図7】図5中の線形伸縮の例を示す図。

【図8】図1中の重複加算合成部の具体的処理手順の例を示す図。

【図9】ガウス窓とハミング窓とを用いた短時間スペクトルの各例を示す図。

【図10】実験結果を示す図。

【図1】

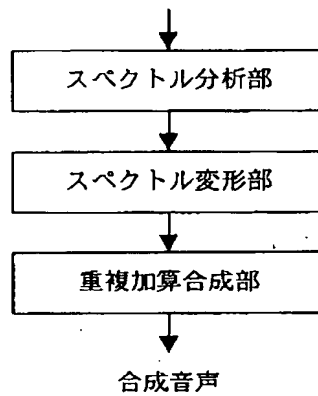


図 1

S0201

S0202

S0203

【図3】

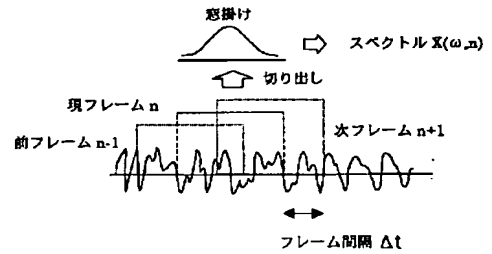


図 3

【図6】

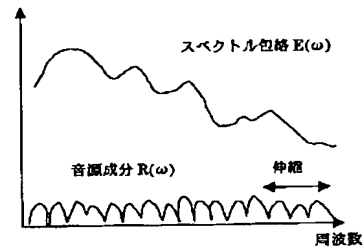


図 6

【図7】

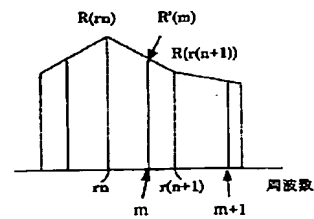
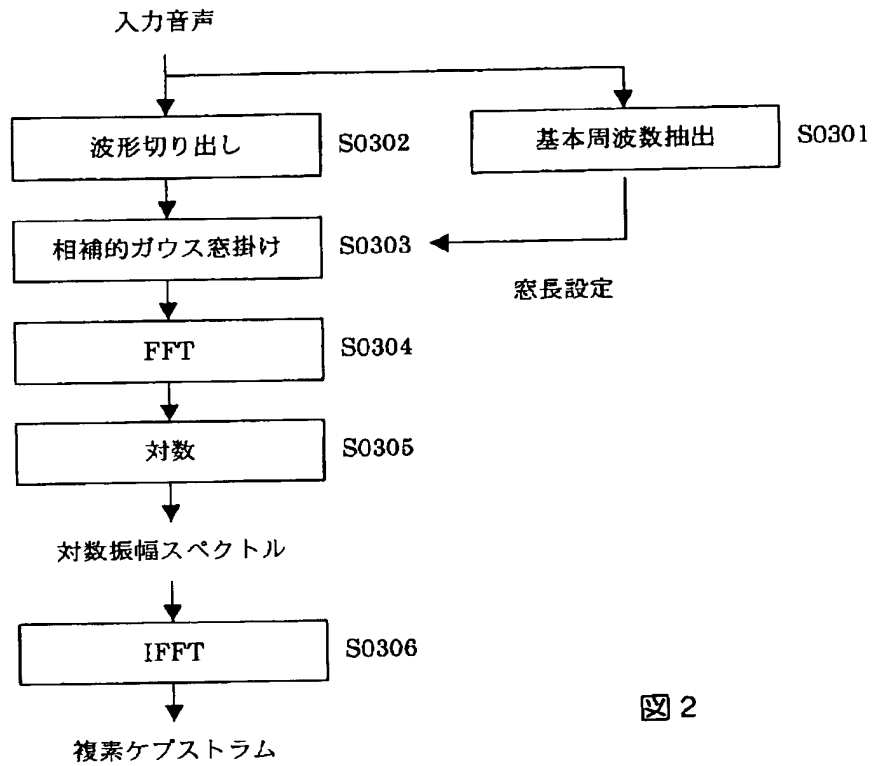


図 7

【図 2】



【図 4】

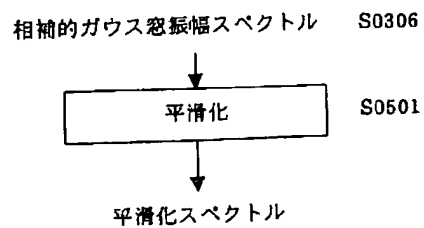
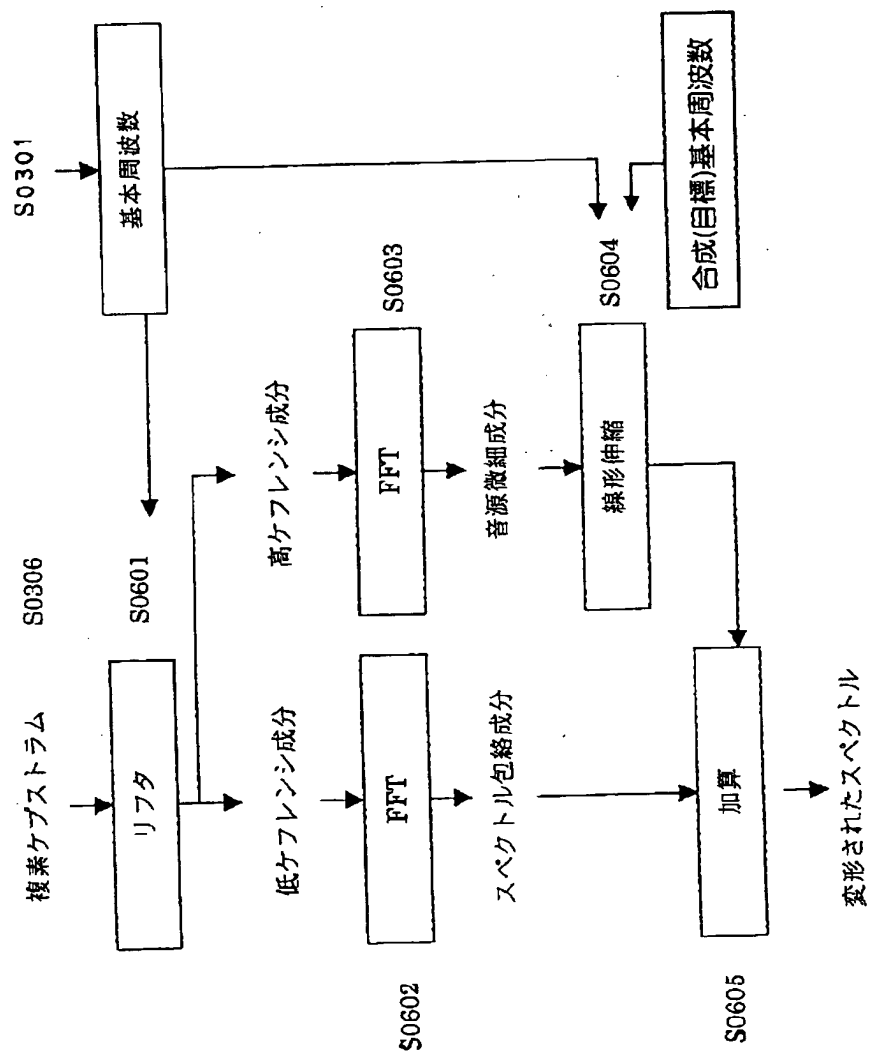


図 4

5
X



【図 8】

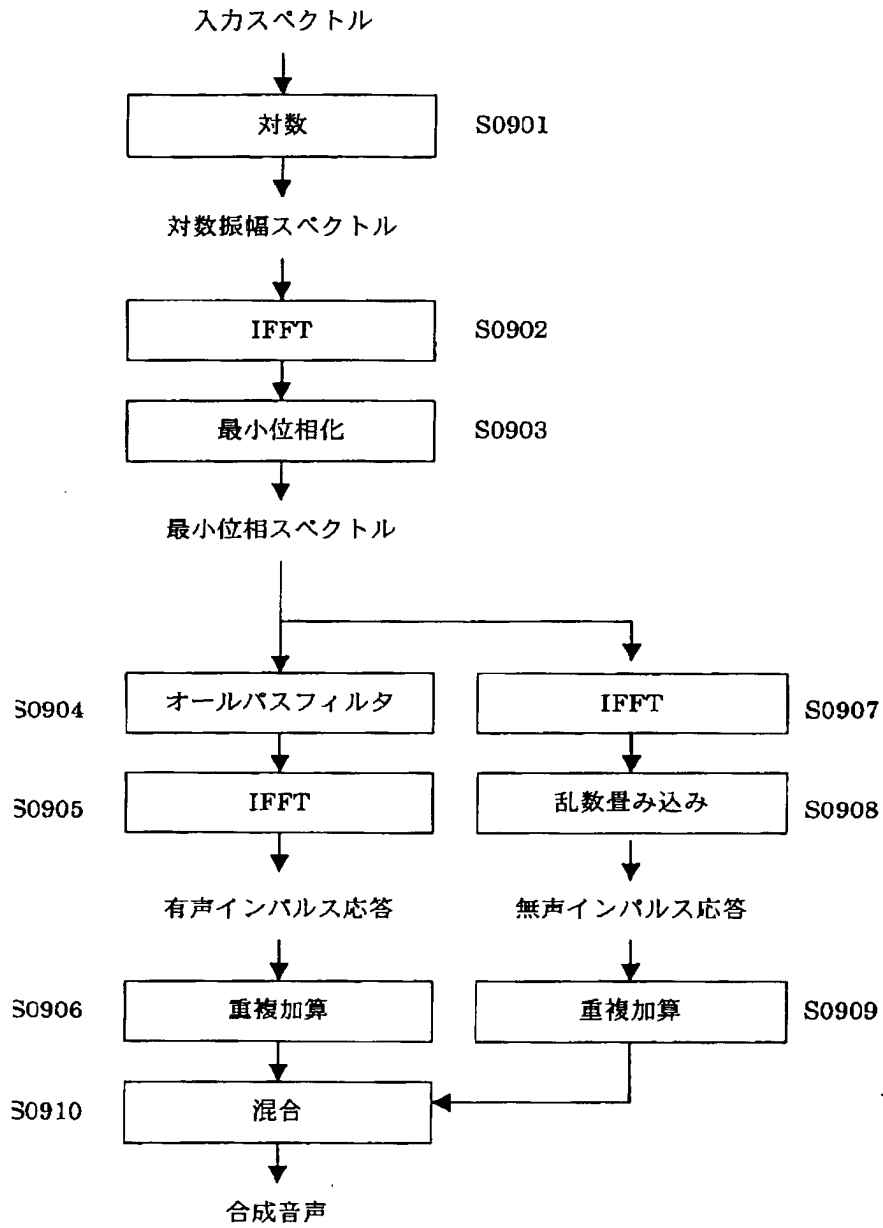


図 8

【図 9】

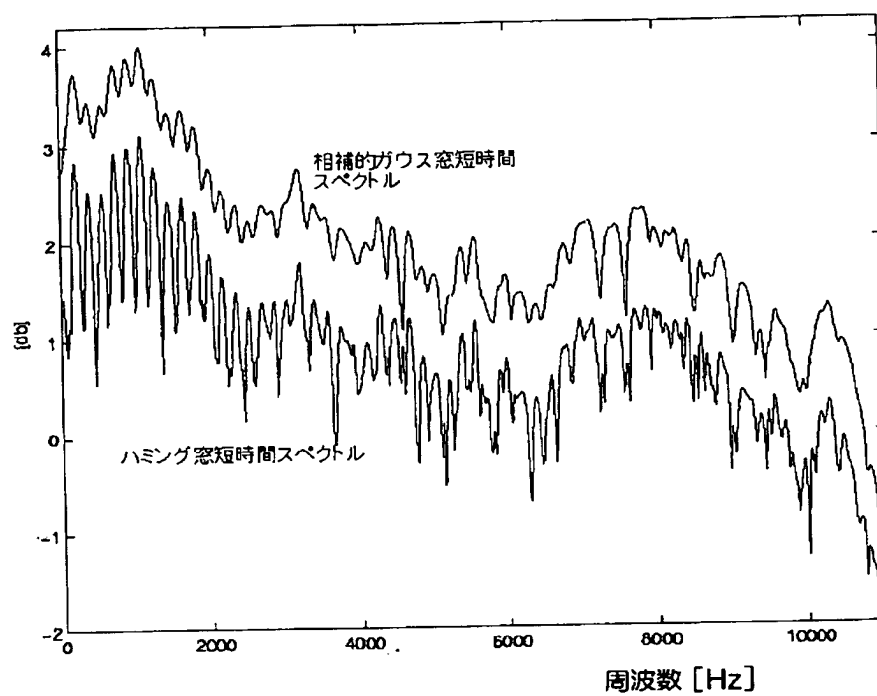


図 9

【図 10】

